✳ **SCIENCE AND CULTURE**

Musical genes

John Carey, *Science Writer*

As an undergraduate at the University of California, Los Angeles (UCLA) in the mid-2000s, Rie Takahashi became so enthralled by science that she shelved her hopes of becoming a professional pianist. But she did have mixed feelings. "I was torn," the contest-winning musician recalls. "I grew up with classical music and piano, and wanted it to be an active part of my life. So I thought, wouldn't it be great to tie these two together?"

Takahashi soon got her chance. In an honors science and society seminar in the mid-2000s, molecular genetics professor Jeffrey H. Miller challenged his students to work on projects that might improve or contribute to society. "I suggested to her that, since she had this unique ability, she tackle the problem of putting DNA sequences to music," says Miller.

Takahashi's wasn't the first effort to turn scientific information into musical notes (see, for example, bit.ly/1V1gQJJ). In the late 1970s, for example, jazz French horn player and Yale professor Willie Ruff and Yale geologist John Rodgers created the "music of the spheres" by turning the movement of the planets into melodies. Recent attempts in the life sciences to create music from scientific data have included compositions based on gene expression and epigenetics. Some researchers are seeking to craft pleasing musical arrangements. But there are aspirations for practical aims as well, such as spotting differences in gene sequences or helping the sight-impaired decipher data.

Key Innovations

Takahashi wanted to create something more musically satisfying than work that had come before, as well as helping to

spot genetic differences. Assigning a note to each of the four nucleotides in a DNA sequence obviously creates a four-note melody, very boring. And representing each of the 20 amino acids that make up proteins with a single note, thus creating a 20-note scale, isn't much better. "The range is enormous," says Miller. "It just sounds like alien music."

So Takahashi came up with two key innovations. The music would be more interesting if each protein were represented by a chord rather than a single note, she figured. Moreover, the 20-note scale could be cut down to 13 notes by assigning similar amino acids the same chord. The two members of each amino acid pair could still be differentiated by using the basic root chord for one, and the first inversion (where the bottom note is flipped to the top of the chord) for the other. For example, the root note for both tyrosine and phenylalanine was G. But Takahashi moved the root of the chord (i.e., the G) for phenylalanine so that it "sounds" different from tyrosine. "She solved the problem of sequence data leading to boring or alien-sounding 'music,' a problem that people had been trying to solve for years, in a really neat and scientifically honest way," says Miller.

Takahashi and Miller worked with UCLA computational biologist Frank Pettit to create a computer algorithm, dubbed Gene2Music, to turn the protein sequences into notes. At first, though, she laboriously did the translations by hand. The first few lines of music, using the amino acid sequence from a protein called thymidylate synthase A, were a revelation. "I thought, wow, this is a more lush musical sound than had been heard with the DNA sequences," Takahashi recalls. "We were really surprised that it sounded great." Samples of the music can be heard on the Gene2Music website (bit.ly/1LeF0Lf).

The paper Takahashi and Miller published in *Genome Biology* in 2007 struck a metaphorical chord with both researchers and the public (1). Wayne Grody, professor of pathology and laboratory medicine, pediatrics, and human genetics at UCLA, now regularly shows his first year medical students slides with Gene2Music renditions of both thymidylate synthase and the mutated huntingtin protein that causes Huntington's disease; he views it as an entertaining educational tool. The difference is striking, because the mutant huntingtin has numerous repeats of three DNA building blocks (cytosine, adenine, and guanine) that appear as the same chord bar after bar. "It always gets a good laugh, and I hope also helps to drive home that what is 'boring' in music is quite disruptive for a protein," says Grody. Adds Miller, "It's very powerful, because music is really very powerful."

Huntingtin
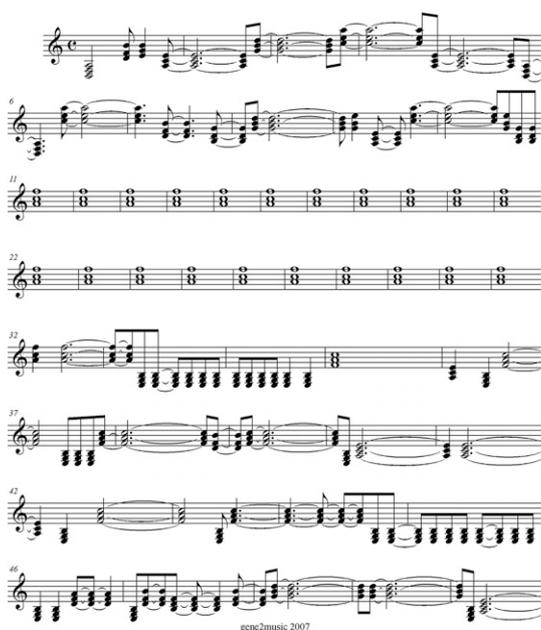
gene2music 2007

The tune above is based on the protein responsible for Huntington's disease. The notes differ drastically from the "normal" case, reflecting how a repeated glutamine sequence causes the protein to malfunction. Listen to the music clip here: bit.ly/1op2J6H. Image courtesy of Rie Takahashi and Jeffrey H. Miller (University of California, Los Angeles, CA).

Musical Expression

That impact can be seen in gene activity and epigenetics as well. At a dinner at a 2014 conference, David Brocks, a graduate student in the division of epigenomics and cancer risk factors at the German Cancer Research Center in Heidelberg, the discussion turned to translating the genetic code to music. Why not, he thought, do the same with epigenetic data?

Brocks set to work. On a single DNA molecule, each site where a cytosine nucleotide occurs next to a guanine nucleotide (called a CpG dinucleotide) can have, or not have, a chemical entity called a methyl group attached. In general, the presence of methyl groups can be associated with turning genes off. Embryonic stem cells have low levels of methylation, whereas differentiated cells have higher levels. And methylation patterns often differ among normal and cancerous cells.

But there was a musical challenge. A CpG dinucleotide is either methylated or it's not, offering up only two possible notes. So Brocks divided those sites into groups of seven, then assigned each of the 128 possible combinations of 0s and 1s to different notes, teaching himself Java along the way to do the translation. The first results were "horrible," he says. "So I dug into melody and harmony and some music theory to change the notes to make it more appealing." Brocks limited his compositions to 10 chords, each with two inversions and four different lengths, and also added three quick extra notes when the DNA sequence had only one methylated site of the seven.

The resulting "music," published in *Clinical Epigenetics* in 2015, sounds a bit like a pleasing piano exercise designed for an intermediate student (2). Perhaps more important, the epigenetic differences between embryonic stem cells and adult cells, or between normal genes and those from prostate cancer cells, are immediately obvious to the ear. That finding suggests music might actually help scientists spot epigenetic patterns, Brocks says.

The ability of music to vividly capture epigenetic differences piqued the interest of Jeffrey Craig, Group Leader for Early Life Epigenetics at the University of Melbourne and Deputy Director of the Australian Twin Registry. A few years ago, a pair of twins asked Craig if he could turn their genome data into music; he's been thinking about the prospect ever since. Identical twins, he notes, can be very different epigenetically, occasionally even more so than two unrelated individuals. With the aid of Brocks' technique, twins, Craig reasons, might be able to hear—and express—those differences musically, helping twins to forge their own individual identities. "It would be a great way to get twins talking to researchers," he adds, noting the need for twin-study subjects. Craig is looking for funding sources to advance such an effort, possibly from an arts organization.

Audible Ambitions

Takahashi suggests that the main scientific goal of such work is to "use the music to recognize patterns that might not be obvious from looking at long sequences." For example, cancer biologist Martin Staege at Martin Luther University Halle-Wittenberg in Germany recently turned microarray gene-expression data from cancer cells into music with a complex algorithm he created called GEMusicA. The algorithm not only assigns notes and note duration to the data, but can also merge the results with famous melodies, like Wagner's "Ride of the Valkyries." The compositions make the discrepancies between neuroblastoma cell lines and the Ewing sarcoma cell line "obvious," Staege reports in a paper published last October (3). He sees music as a tool for the analysis of differential gene expression, though small-scale tests with 23 subjects suggested that music is not superior to visual study as a means of discriminating gene expression. Musical transformations are perhaps more likely to remain an entertaining curiosity, especially given how expert computers have become at discerning differences.

Explaining science to young people is difficult, but music is universal.

—Jeffrey Miller

Even so, the ability to "hear" genetic or epigenetic data could serve as a teaching tool, making the science more accessible, the researchers say. "Explaining science to young people is difficult, but music is universal," says Miller. "It's a great teaching tool."

Now an MD-PhD and a dermatology medical resident studying stem cells and skin cancer, Takahashi continues to create genetic compositions despite a busy schedule and no funding. In 2012, for example, she published a paper in *Autophagy* with yeast biologist Daniel Klionsky of the University of Michigan (4). The authors describe turning autophagy, an internal cell process in which cell components are transported to lysosomes for degradation and recycling, into a musical piece they call the "SNARE Dance" using the Gene2Music program (4). The autophagy example they picked involves four proteins, each represented by a different instrument: marimba, flute, harp, and a Japanese lute called a shamisen. "Each protein has its own personality and its own musical line," says Takahashi. "Imagining all these coming together was an adventure."

Her musical accomplishments have helped Takahashi both connect with nonscientists and stand out from her fellow students and researchers, she says. When she interviewed for MD-PhD and residency positions, "nine times out of ten, the interviewers wanted to talk about my music, not my wet lab research," Takahashi says with a laugh. "Some even had the music playing."

1 Takahashi R, Miller JH (2007) Conversion of amino-acid sequence in proteins to classical music: Search for auditory patterns. *Genome Biol* 8(5):405.

2 Brocks D (2015) Musical patterns for comparative epigenomics. *Clin Epigenetics* 7(1):94.

3 Staege MS (2015) A short treatise concerning a musical approach for the interpretation of gene expression data. *Sci Rep* 5:15281.

4 Takahashi R, Miller J, Klionsky DJ (2012) SNARE Dance: A musical interpretation of Atg9 transport to the tubulovesicular cluster. *Autophagy* 8(3):294–296.